Standardized Incidence Ratio (SIR)

A Math-based Approach to Evaluating Unusual Patterns of Cancer

People concerned about unusual patterns of cancer in their community often ask their state health department for an evaluation. CDC/ATSDR provide guidance to assist state health departments when addressing such community concerns (see 2022 Guidelines for Examining Unusual Patterns of Cancer and Environmental Concerns). One of the methods to evaluate if cancer is occurring at an unusual rate involves calculating a standardized incidence ratio (SIR).

What is a standardized incidence ratio (SIR)?

The SIR is an estimate of the number of cancer cases in a given population compared to what might be "expected" based on a comparison with the cancer experience in a larger population. Thus, the SIR is a ratio of the number of cancers observed compared to the number expected. If the observed number of cases equals the expected number of cases, the value of an SIR is 1.0. Sometimes we multiply the SIR by 100 for ease of reporting.

What does it mean for the incidence ratio to be standardized?

Standardizing is a way to adjust a rate by taking into account factors about a population. For example, rates can be adjusted for factors such as age, sex, race, or ethnicity. In cancer analyses, adjusting for age is important because age is a risk factor for many cancers. Therefore, if the comparison population is composed of an older population, you might expect cancer to occur more frequently because cancer rates increase with age. In this example, the calculation of the SIR would take the age of the population into account and adjusts the comparison based upon the age of that population. We call this standardizing by age.

If the SIR is greater than 1.0 (or 100 if multiplied by 100), should I be concerned?

Not necessarily; as a statistical estimate, the interpretation of the SIR is not straightforward. SIRs that are greater than 1.0 (or 100) should be explored further.

Sometimes an SIR greater than 1.0 (or 100) may be due to random chance. For example, if you divide a county into 10 squares with 200 people living in each, you would see some squares with no cases, while other squares would have multiple cases, just by chance alone.

Similarly, if the number of cancer cases is small, that can dramatically impact the SIR. Small numbers can make annual changes in the rate seem large even though the average over time remains the same as the expected number. Small numbers make it even more important to do precise counts. If a single case is missed or counted twice, it could make a big difference (and bias the results away from the truth).

What is a confidence interval?

The confidence interval (CI) is the range of values within which the true values lie. The calculation of the CI helps to describe the stability and precision of the calculated SIR. The wider the confidence interval, the more variability there is and the less confidence you have that the estimated SIR is the true SIR. The larger the numbers of observed and expected cases in the population, the more likely you are to have a stable range of values, within which the true SIR lies; thus, a more precise confidence interval. For example, a 95% confidence interval would mean that we are 95% confident that the true value falls somewhere in that range.

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

CS 335483-A 11/16/2022

How is an SIR calculated?

Using an example from CDC's Environmental Public Health Tracking Network for Bingham, Idaho, from 2014-2018 there were 36 cases of leukemia. In 2016, there were 45,313 people living in this county, so the **cancer rate** was around:

$$\frac{\text{36 cases}}{\text{45,313 people each year}} = 0.0008$$

Based on cancer rates for the state, if the expected number of cases was 38, we would say the actual number of cases was lower than the expected number of cases. The **SIR** would be:

$$\frac{\text{36 cases per year (actual)}}{\text{38 cases per year (expected)}} = 0.96 \text{ (x 100)}$$

Now if the number of actual cases in the community was 38, we would say that the number was equal to the expected number of cases. In that case 38 would be divided by 38 to get an SIR equal to 1.0 (x 100).

If the number of actual cases was 76, then you would divide 76 by 38 to get an SIR of 2.0 (x 100). In this case, we would say that the actual number of cases was greater than expected.

In this example, these calculations involve over-simplifications, ignoring year-to-year variability and by not addressing the age adjustment that accounts for differing age distributions in different populations.

When looking at an SIR, one should consider not just its value, but also how close that value is to 1 (or 100), as well as how wide the confidence interval is.

How is a confidence interval calculated?

As mentioned above, we cannot always be certain that the calculated number represents the true value. The **confidence interval** is a range of numbers above and below the SIR. A 95% confidence interval would mean that we are 95% confident that the true value falls somewhere in that range.

If the confidence interval is wide, it shows a lot of uncertainty around the value. This is usually based on the number of cases. For example, an SIR of 2.0 (or 200) could be based on a ratio of observed to expected cases of 10/5, or 1000/500, but the confidence interval based on 1000/500 would be narrower. Having a narrower confidence interval increases confidence in our estimate.

Applying the equation for calculating a confidence interval provided in the 2022 Guidelines, for the Bingham, Idaho example discussed above (where there were 36 observed cases and 38 expected), the confidence interval (CI) ranged from 0.67 to 1.29. We would write this as an SIR of 0.96 (95% CI: 0.67, 1.29).

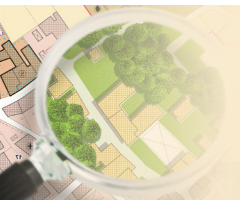
More information about confidence intervals can be found at Statistical Methods: Confidence Intervals | U.S. Cancer Statistics Data Visualizations Tool Technical Notes | CDC.

Is there a threshold of SIR values above which there may be more cause for concern?

Various factors associated with the calculation of SIRs complicate the interpretation of the values, and applying such rigid values around the interpretation of SIRs would discount other important information such as:

- the history of the community, especially as it relates to environmental exposures,
- the potential to undercount cases because of the long time that typically occurs between exposures and cancer onset, and because people move in and out of the community,
- the fact that small numbers of rare cancers may warrant more careful review, and
- the confidence interval for the SIR.

For this reason, CDC does not recommend applying threshold values.



For more information, see Unusual Patterns of Cancer, the Environment, and Community Concerns: https://www.cdc.gov/nceh/cancer-environment/index.html